DOCUMENT RESUME

ED 231 313 HE 016 382

AUTHOR Aleamoni, Lawrence M.; Stevens, Joseph J.
TITLE The Effectiveness of Consultation in Support of Student Evaluation Feedback: A Ten Year Follow-up.
PUB DATE 83
NOTE 21p.; Paper presented at the Annual Meeting of the Rocky Mountain Psychological Association (Albuquerque, NM, Apr. 28-May 1, 1983).
PUB TYPE Reports - Research/Technical (143) -- Speeches/Conference Papers (150)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS *College Faculty; *Consultation Programs; *Faculty Development; *Feedback; Followup Studies; Higher Education; *Instructional Improvement; *Student Evaluation of Teacher Performance; Teacher Effectiveness; Teacher Workshops

ABSTRACT

A 10-year followup was conducted of a study that examined the effectiveness of expert consultation and student feedback provided to college faculty. A 1978 study (Aleamoni) found significant improvement in student ratings for a group of instructors receiving student feedback and expert consultation in comparison to a group of instructors receiving student feedback alone. An assessment was undertaken with 28 of the 33 instructors who participated in the original 1971 study and had voluntarily used the Course/Instructor Evaluation Questionnaire after the termination of the original study. The number of times instructors had engaged in individual consultations or had participated in instructional workshops during the followup period was also assessed. It was found that instructors who had participated in consultation received higher student ratings and used student ratings and instructional services more throughout the followup interval. Despite methodological difficulties in the research, it is concluded that the usefulness of student rating feedback is ensured only when integrated with a system of instructional support. (SW)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED231313

The Effectiveness of Consultation in Support of Student Evaluation Feedback: A Ten Year Follow-up

Lawrence M. Aleamoni and Joseph J. Stevens

University of Arizona

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

Joseph J. Stevens

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☐ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

AE016 382

Abstract

A study by Aleamoni (1978) found significant improvement in student ratings for a group of instructors receiving student feedback and expert consultation in comparison to a group of instructors receiving student feedback alone. The present study was a follow-up, after ten years, of the instructors who participated in the original study. It was found that instructors who had participated in consultation received higher student ratings and used student ratings and instructional services more throughout the follow-up interval. While methodological difficulties were present in the study, it was concluded that the usefulness of student rating feedback is ensured only when integrated with a system of instructional support. A critical aspect of the instructional support system involves training the instructor how to effect instructional improvement.

The Effectiveness of Consultation in Support of Student Evaluation Feedback: A Ten Year Follow-up

The use of student ratings as a method of course and instructor evaluation has increased substantially over the past ten years. With this increase a number of standardized instruments for assessing instructional effectiveness have become available. These instruments provide a reliable and relatively simple methodology for obtaining student evaluations of instruction. Recent reviews of the validity of ratings have established the usefulness of student ratings as a measure of instructional effectiveness (Aleamoni, 1980; Centra, 1979; Cohen, 1981; Kulik & Kulik, 1974; Marsh, 1980; McKeachie, 1979; Millman, 1981). Information derived from student evaluations, however, may serve a number of purposes. Cohen (1980) defined three such purposes: (1) to aid in administrative decisions, (2) to aid students in course/instructor selection, and (3) to provide feedback to instructors for instructional improvement.

This last purpose of student evaluations, that of instructional improvement, was the focus of the present paper. Some disagreement exists in recent reviews regarding the effectiveness of student evaluations for improving instruction. A review by Rotem and Glasman (1979) concluded that "feedback from student ratings . . . . does not seem to be effective for the purpose of improving performance of university teachers" (p. 507). However, a number of studies have found substantial increases in student ratings as a function of feedback to the instructor. For example, a study

by Overall and Marsh (1979) found that, after feedback to the instructor, not only did student ratings increase, but also student achievement and motivation.

The discrepancies in reported results may be attributed to differences in instrumentation and in methodology or factors which reduce the efficacy of feedback. A number of authors have specified factors which may be responsible for failures of instructor improvement after student-ratings feedback (Cohen, 1980; Kulik & Kulik, 1974; McKeachie, 1979). No instructional improvement may occur when the feedback does not provide new information to the instructor. Additionally, instructor improvement may be a function of both time allowed for change and the instructor's self-rating of instructional effectiveness. That is, changes may not be evidenced within a single semester. And, the instructor's willingness to accept evaluation may be lowered if the evaluation is inconsistent with self-perceptions. Lastly, and perhaps most importantly, the instructor may not know <u>how</u> to change.

The effectiveness of student evaluations for instructional improvement was reviewed most recently by Cohen (1980). Cohen conducted a meta-analysis of instructional feedback studies. From the analysis it was concluded that feedback had a modest but significant effect (15 percentile points) in improving instruction. Cohen also found that this effect was accentuated when consultation accompanied feedback. The purpose of this paper is to suggest that expert consultation may ameliorate many hypothesized factors which inhibit the effectiveness of instructional feedback. That is,

instructional development and improvement is facilitated when an accessible system of instructional support is available to the instructor. Without such an instructional system, it is hypothesized that student ratings feedback will produce moderate but somewhat inconsistent gains in instructional improvement.

One study which demonstrated an increase in student ratings after feedback and consultation to instructors was conducted by Aleamoni (1978). Student rating feedback was provided by administration of the third generation of the Course/Instructor Evaluation Questionnaire (CIEQ). In addition, twenty of the thirty-three instructors participating in the study engaged in individual diagnostic consultations with the author. After a one-semester delay, CIEQ student ratings were found to be significantly higher for those instructors who had received feedback and consultation in comparison to instructors receiving feedback (CIEQ results) only.

The original study arose as a result of interactions in a three-day workshop held at the University of Arizona during academic year 1971-72. In 1975, three years after the conclusion of the study, the University of Arizona founded the Office of Instructional Research and Development (IRAD). Previously, no systematic source of instructional support was available. IRAD has since offered a number of services to instructors including: (1) course/instructor evaluation using the CIEQ, (2) an Optional Item Catalog for more individualized evaluation (Aleamoni & Carynnyk, 1977), (3) interpretation manuals to aid in normative decile comparison, problem identification, and diagnosis using CIEQ (Aleamoni, 1979), (4) "open-door" individual consultation with instructors, and (5) regularly scheduled workshops on methods of improving instruction and instructional effectiveness.

The establishment of the office of IRAD after the conclusion of the original study resulted in a unique opportunity to examine the long-term effects of instructional intervention. The combination of individual diagnostic consultation with CIEQ feedback resulted in significant improvement for instructors who previously had little access to instructional support. After the establishment of IRAD, instructors who had participated in the original study had the opportunity to voluntarily continue use of the CIEQ and seek instructional support services. The purpose of the present study was to determine, after a ten year interval, the relative standing of the original experimental groups. Additionally, an attempt was made to infer whether voluntary usage of instructional services was different for the two groups during the ten year interval.

Method

Sample

In the original study 33 instructors at the University of Arizona who had used the CIEQ during the Fall 1971 and Spring 1972 terms were used as subjects. All subjects had voluntarily scheduled to participate in individual diagnostic consultations with the author. However, due to time constraints, 13 of the instructors were not able to participate in consultations. As a result, two groups of subjects were defined: a group which received feedback and consultation (FC) and a group which received feedback only (FO).

The FC group consisted of 17 instructors teaching 24 courses. The FO group consisted of 13 instructors teaching 18 courses. In the original

study, the "experimental" group also included 8 instructors teaching 10 other courses. These courses were systematically excluded from statistical analysis in the original study. The criterion used for exclusion was an obtained normative decile rating of 8 or above on the CIEQ. The original study reported no data for this group; here defined as the excluded experimental (EE) group. It should be noted that the EE group received both feedback and consultation, but no difficulties were identified in the diagnostic consultation.

The present paper is concerned with 28 of the 33 instructors (85%) who had participated in the original study and had voluntarily used the CIEQ after the termination of the original study. Of these instructors, 13 had previously been categorized in the FC group, 7 in the FO group, and 8 in the EE group.

Procedure

For each instructor who had participated in the original study, data were collated for all uses of the CIEQ from Fall term 1975 to Spring term 1982. CIEQ data were included only if the instructor was teaching the same course that had been evaluated in the original study. For the 13 FC group instructors this resulted in evaluations from 74 courses; while 27 course evaluations were obtained for the 7 FO instructors, and 31 for the EE group instructors. Due to the small n-sizes and variations in n-size at each semester interval, the data were collapsed into two five-year follow-up intervals. This resulted in two intervals each of seven consecutive semesters from Fall 1975 to Fall 1978 and from Spring 1979 to Spring 1982. For the first follow-up period there were 35 courses for the FC

group, 15 for the FO group and 15 for the EE group resulting in a total of 65 courses. The second follow-up period contained 39 FC group courses, 12 FO group courses, and 16 EE group courses for a total of 67 courses.

Results

Means and standard deviations were computed for each instructor/course at each of the two follow-up intervals on each of the five subscales of the CIEQ and the total. The data were then analyzed using an unweighted means analysis of variance. Additionally, the mean number of usages of the CIEQ was calculated for each group. Records were also searched to determine the number of times instructors in the FC or FO groups had engaged in individual consultations or had participated in instructional workshops during the follow-up period.

None of the analyses of variance comparing mean ratings on the CIEQ produced significant differences at the .05 level. However, four of the subscales and the total approached significance. The analysis of variance on the difference between groups on the method subscale produced the F ratio with the lowest probability ($F = 3.35$, $df = 1/31$, $MS_E = 86.98$, $P < .08$). The mean ratings on three of the subscales and the total rating of the CIEQ are presented in Figure 1. The pre and post data presented in Figure 1 are the means obtained in the original study by the instructors present during the follow-up.

Insert Figure 1 about here

Inspection of Figure 1 revealed that the trends for the follow-up sample during the original study were essentially equivalent to the results

reported for the entire original sample. The instructors receiving consultation (FC) displayed marked gains in evaluation ratings from the pretest to posttest period. In the original study significant differences were found from pretest to posttest for the FC group. It can also be seen that the FC and FO groups were approximately equal in pretest ratings before the exclusion of EE group instructors. That is, collapsing of the FC and EE group pretest means results in means comparable to those of the FO group prior to the intervention.

Data points in the panels of Figure 1 were not connected between the original study and the follow-up study. As a result of the establishment of IRAD in the interim, substantial and unknown changes occurred between the posttest and follow-up intervals. For example, average ratings lowered regardless of group following the original study. This effect can be attributed to changes in sensitivity or reactivity to evaluation or changes in the student population.

The relative position of groups throughout the follow-up interval, however, was remarkably similar to the posttest results. For example, the difference between the FC and FO groups on the CIEQ Total rating during posttest (.65) increased slightly throughout the ten year follow-up interval (.72 at 5 years, .93 at 10 years). The only subscale that demonstrated a shift in the relative standing of the FC and FO groups was the instructor subscale. The difference in mean ratings on the instructor subscale was small during the posttest (.24) and remained small during the follow-up (.37 at 5 years, .13 at 10 years). Rated evaluations

of the EE group instructors, which were markedly higher than FC group instructors in the original study, were not reliably distinguishable from ratings for the other two groups.

The average number of usages of the CIEQ by the three groups of instructors was different throughout the ten year period. An analysis of variance indicated that this difference was significant ($F = 3.41$, $df = 2/41$, $MS_E = 0.2868$, $p < .04$). The CIEQ was used most by the FC group and least by the FO group. It was also found that CIEQ usage for the FC group increased over the two follow-up intervals and usage for the FO group decreased, though this result was not significant.

Insert Figure 2 about here

Inspection of the records for workshop attendance and participation in individual consulting also demonstrated a difference between the FC and the FO groups. These records cannot be considered exhaustive nor perfectly accurate and were therefore not subjected to statistical test. It is interesting to note, however, the substantially greater frequency of consulting contact evidenced by the FC group of instructors.

Insert Figure 3 about here

Discussion

The results of the present study must be interpreted with caution. Two methodological difficulties were present. Experimental control over the assignment of subjects to the FC and FO groups was not possible in the original study. As a result, no assessment of reactivity to feedback or the relative effects of feedback can be made through the use of a no-feedback control group. Secondly, the lack of random assignments to groups clouds the interpretation of group differences. The mechanism which defined groups in the original study was "schedule conflicts" and could perhaps be considered a "chance" process. Furthermore, Cohen (1980) found no relationship ($r = -.06$) between random assignment of comparison groups and obtained effect size in a meta-analysis of feedback studies. Even so, interpretation of the present study is hampered by inadequate experimental control.

However, substantive conclusions were implied. A number of studies have indicated that expert consultation facilitates the effectiveness of student evaluation feedback. The results of the original study support this conclusion. Furthermore, the follow-up results indicated that instructional intervention may produce long-term effects in instructional effectiveness and instructor behavior.

Student ratings of the FC and the EE group instructors were quite similar after the ten year interval. Student ratings of the FO group instructors, nowever, were remarkably lower on four of the five CIEQ subscales. This result can be interpreted in three ways: (1) the original intervention produced a stable long-term difference between groups, (2) the groups were

inherently different, and (3) the original intervention produced a short-term improvement, but also altered the instructors' strategies for instructional change.

It is hypothesized that the last interpretation provides the best explanation of the obtained results. Instructors who participated in the original FC group subsequently made use of student evaluation feedback and instructional services more frequently than FO group instructors. The original diagnostic consultation may have resulted in a generalization by the instructor of how to pursue instructional improvement. As a result, the instructor was able to effect change in the specific course of interest. But the instructor was also exposed to a general strategy for effecting instructional change. This strategy would include knowledge of how to analyze and interpret feedback information and knowledge of how to seek instructional support when needed. The occurrence of such a generalization would result in a greater likelihood of the usage of both evaluational feedback and instructional services by the instructor. This interpretation was supported in the present study by the significantly higher usage of the CIEQ and the greater frequency of usage of instructional services by the FC group instructors over the ten year follow-up period.

As Rotem and Glasman (1979) point out, the effectiveness of student evaluations is dependent on their use and intended purpose. But when student evaluation is used to provide feedback for purposes of instructional improvement and development, it is unlikely that feedback of results alone will suffice. For example, Cohen (1980) found that the provision of normative comparisons of feedback results was not related to instructional

improvement. The unwarranted assumption is that the instructor knows how to make use of such comparisons. This assumption is justified only when the instructor is trained to use feedback information effectively. That is, information can be used for improvement and development only if the instructor knows how to interpret and apply the information.

The provision of evaluational feedback is but one aspect of improving instructional effectiveness. Consistent instructional improvement is also dependent on the availability of a coherent system of instructional resources. Without such a system, the instructor may be unable to gain either the knowledge or the professional support necessary to effect change. In the present study, the FC group instructors utilized such resources more frequently than the FO group instructors. Over a ten year interval, the FC group instructors also consistently received higher ratings than the FO group instructors. Given the descriptive nature of the present study, no strong connection can be assumed. However, the results of the present study imply that integration of the instructor with a system of instructional resources produces marked long-term differences in the usage of services and rated teaching effectiveness. In any event, it seems reasonable to conclude that the instructor cannot effectively change, unless the instructor knows how to change.

References

Aleamoni, L. M. The usefulness of student evaluations in improving college teaching. Instructional Science, 1978, 7, 95-105.

Aleamoni, L. M. Arizona course/instructor evaluation questionnaire (CIEQ): Results interpretation manual, form 76. Tucson, Arizona: University of Arizona, Office of Instructional Research and Development, 1979.

Aleamoni, L. M. The use of student evaluations in the improvement of instruction. National Association of Colleges and Teachers of Agriculture Journal, 1980, 24(3), 18-21.

Aleamoni, L. M., & Carynnyk, D. B. Optional item catalog (revised) (IM 6). Tucson, Arizona: University of Arizona, Office of Instructional Research and Development, 1977.

Centra, J. A. Determining faculty effectiveness. San Francisco: Jossey-Bass, 1979.

Cohen, P. A. Effectiveness of student-rating feedback for improving college instruction: A meta-analysis of findings. Research in Higher Education, 1980, 13, 321-341.

Cohen, P. A. Student ratings of instruction and student achievement: A meta-analysis of multisection validity studies. Review of Educational Research, 1981, 51, 281-309.

Kulik, J. A., & Kulik, C.-L. C. Student ratings of instruction. Teaching of Psychology, 1974, 1, 51-57.

Marsh, H. W. Research on students' evaluations of teaching effectiveness: A reply to Vecchio. Instructional Evaluation, 1980, 4, 5-13.

McKeachie, W. J. Student ratings of faculty: A reprise. Academe, 1979, 65, 384-397.

Millman, J. (Ed.). Handbook of Teacher Evaluation. Beverly Hills, California: Sage Publications, 1981.

Overall, J. V., & Marsh, H. W. Midterm feedback from students: Its relationship to instructional improvement and students' cognitive and affective outcomes. Journal of Educational Psychology, 1979, 71, 856-865.

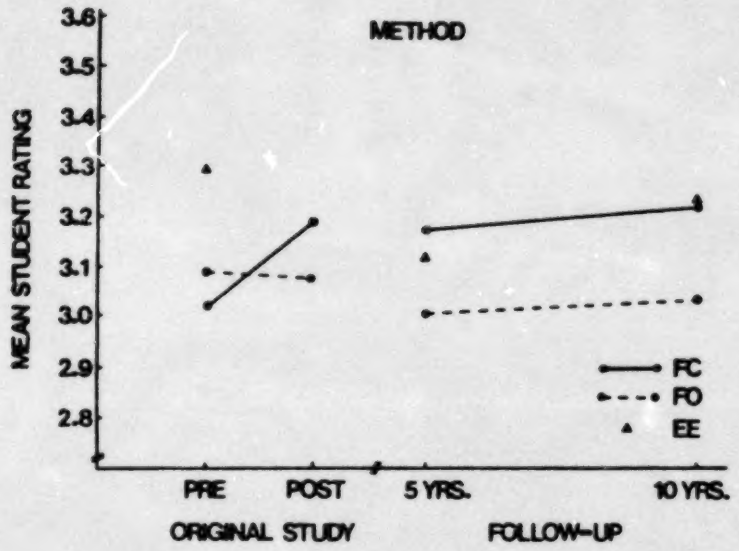
Rotem, A., & Glasman, N. S. On the effectiveness of students' evaluative feedback to university instructors. Review of Educational Research, 1979, 49, 497-511.

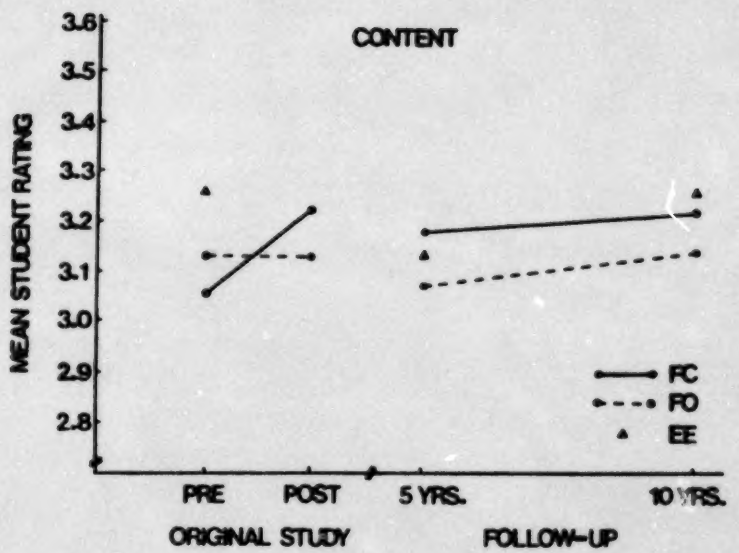
Figure Captions

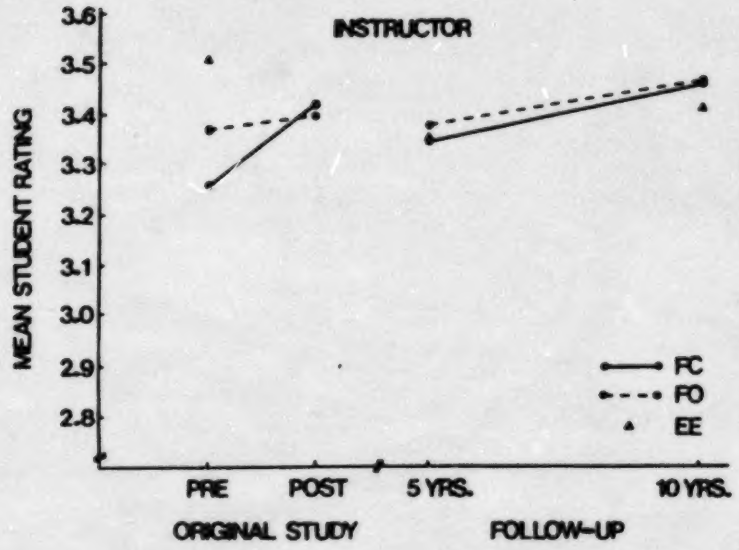
Figure 1. Mean CIEQ ratings by group and subscale over a ten year period.

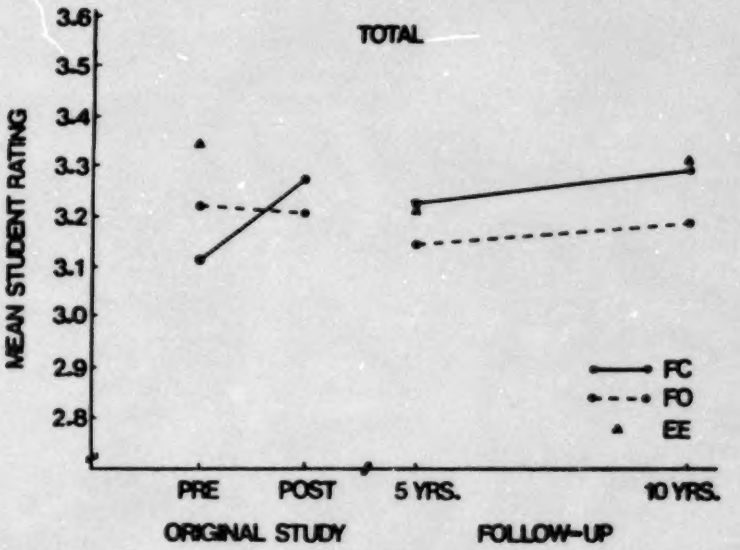
Figure 2. Mean number of uses of the CIEQ by group during the follow-up interval.

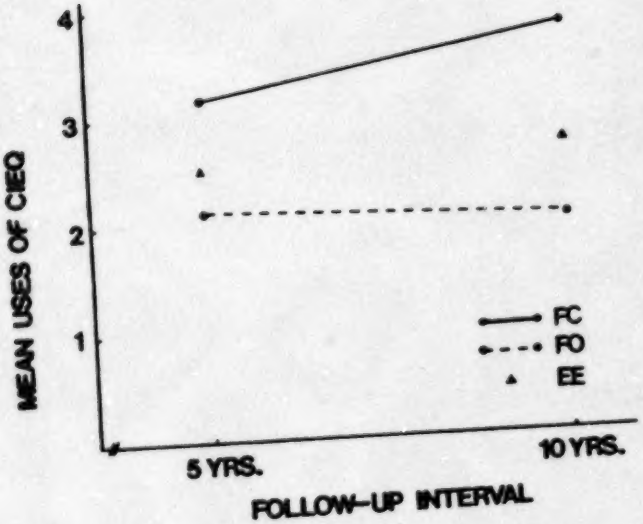
Figure 3. Number of workshop attendances or consultations by group during the follow-up interval.

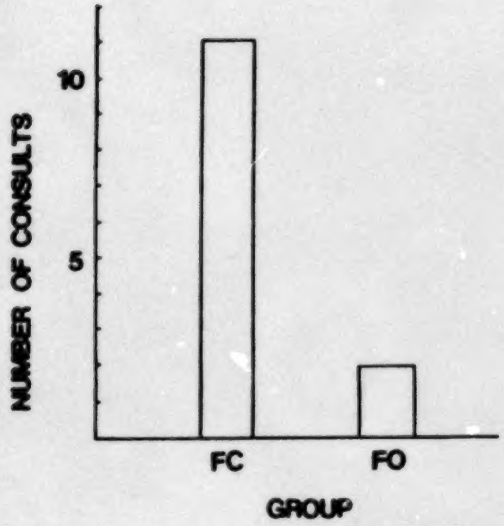

METHOD
MEAN STUDENT RATING
3.6
3.5
3.4
3.3
3.2
3.1
3.0
2.9
2.8
FC
FO
EE
PRE
POST
5 YRS.
10 YRS.
ORIGINAL STUDY
FOLLOW-UP


CONTENT
MEAN STUDENT RATING
3.6
3.5
3.4
3.3
3.2
3.1
3.0
2.9
2.8
FC
FO
EE
PRE
POST
5 YRS.
10 YRS.
ORIGINAL STUDY
FOLLOW-UP


INSTRUCTOR
MEAN STUDENT RATING
3.6
3.5
3.4
3.3
3.2
3.1
3.0
2.9
2.8
FC
FO
EE
PRE
POST
5 YRS.
10 YRS.
ORIGINAL STUDY
FOLLOW-UP


TOTAL
MEAN STUDENT RATING
3.6
3.5
3.4
3.3
3.2
3.1
3.0
2.9
2.8
FC
FO
EE
PRE
POST
5 YRS.
10 YRS.
ORIGINAL STUDY
FOLLOW-UP


MEAN USES OF CIEQ
4
3
2
1
FC
FO
EE
5 YRS.
10 YRS.
FOLLOW-UP INTERVAL


NUMBER OF CONSULTS
10
5
FC
FO
GROUP